

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Nov – Dec 2023

Program: B. Tech Scheme : II

Examination: LY Semester: VII

Course Code: CEC702 Course Name: Big Data Analytics

Date of Exam: 10/12/2023

Duration: 2.5 Hours

Max. Marks: 60

Instructions:

- (1) All questions are compulsory.
- (2) Draw neat diagrams wherever applicable.
- (3) Assume suitable data, if necessary.

		Max. Marks	CO	BT level
Q 1	Solve any six questions out of eight:	12		
i)	List and explain characteristics of big data.	02	CO1	U
ii)	Explain NoSQL business drivers.	02	CO2	U
iii)	Explain any four HIVE commands in brief.	02	CO1	U
iv)	Explain role of combiner in Mapreduce.	02	CO3	U
v)	Explain challenges in Stream Processing.	02	CO4	U
vi)	Explain dead end and spider traps with sample web graph.	02	CO5	U
vii)	Explain how DataFrame is created using R. Demonstrate any two operations on it.	02	CO6	U
viii)	Explain Union operation with Mapreduce.	02	CO3	U
Q.2	Solve any four questions out of six. (M – Marks)	16		
i)	'LinkedIn suggests you to connect with people you may know,' Which type of recommendation is this? Explain it.	04	CO5	An
ii)	Describe Hadoop architecture in detail with its advantages and limitations.	04	CO1	U
iii)	Explain any one algorithm to count number of distinct elements in a Data stream.	04	CO4	U
iv)	Explain grouping and aggregation using MapReduce with example.	04	CO3	U
v)	Explain BASE model with reference to NOSQL.	04	CO2	U
vi)	List and explain data types in R programming. (1 M) How different types of operators are used to perform computations? (3 M)	04	CO6	U
Q.3	Solve any two questions out of three. (M – Marks)	16		
i)	With suitable block diagram explain architecture of HDFS. (3 M) Discuss role of Data node and Name node in HDFS. (2 M)	08	CO2	U

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Nov – Dec 2023
 Program: B. Tech Scheme : II .
 Examination: LY Semester: VII
 Course Code: CEC702 Course Name: Big Data Analytics

Date of Exam: /12/2023 Duration: 2.5 Hours Max. Marks: 60

	Briefly explain handling of node failure.(3 M)			
ii)	What is a Community in a Social Network Graph? (1M) Explain steps of Girvan Newman algorithm. (3 M) Apply Girvan Newman on sample web graph to detect communities.(4 M)	08	CO5	Ap
iii)	Explain any four types of charts in R for data visualization.(2 M) Write R code statements to prepare these charts.(6 M)	08	CO6	U
Q.4	Solve any two questions out of three. (M – Marks)	16		
i)	Explain DGIM algorithm with example. (4 M) How long to continue with bucket formation? (2 M) When to merge two buckets? What will be the timestamp of newly formed bucket? (2 M)	08	CO4	An
ii)	What is NoSQL database? (1 M) List the differences between NoSQL and relational databases. (4 M) Explain in brief various types of NoSQL databases in real time applications. (3 M)	08	CO2	U
iii)	What is PageRank? (2 M) Explain damping factor. (2 M) Show PageRank computation for two iterations for above web graph. (4 M)	08	CO5	Ap


